L'Europe des humanités au cœur du Quartier latin



2013 – 2014

Domaine : Sciences humaines et sociales

Mention : **Information et communication**

Spécialité : **Informatique et ingénierie de la langue pour la gestion de l'information**

Finalité : **Recherche**

PARIS-SORBONNE

lettres et civilisations



2013-2014

Domaine

« Sciences humaines et sociales »

Mention

« Information et Communication »

Spécialité

« Informatique et Ingénierie de la langue
pour la Gestion de l'Information »

Finalité

Recherche

Sommaire

UFR Sociologie et informatique pour les sciences humaines

Directeur: Pierre DEMEULENAERE

Département d'Informatique, Mathématiques et de Linguistique appliquées

SECRETARIAT DU MASTER

Monsieur Joseph Chauleau
Maison de la Recherche, 28 rue Serpente – 75006 Paris
Téléphone : 01.53.10.58.31
Email : joseph.chauleau@paris-sorbonne.fr
Ouverture de 9h30 à 12h30 et de 14h00 à 16h30

Les enseignements ont lieu à la Maison de la Recherche

Equipe enseignante :

Professeurs

Caraty Marie-José	Professeur à l'Université Paris V (27ème section)
Devillers Laurence	Professeur à l'Université Paris IV (27ème section)
Ibrahim Amr Helmy	Professeur à l'Université de Franche-Comté (7ème section)
Montacié Claude	Professeur à l'Université Paris IV (27ème section)

Maîtres de conférences

Eyharabide Victoria	Maître de Conférences à l'Université Paris IV (27ème section)
Gorgatchev Bissera	Maître de Conférences à l'Université Paris XIII (7ème section)
Guérin Françoise	Maître de Conférences à l'Université Paris IV (7ème section)
Jackiewicz Agata	Maître de Conférences à l'Université Paris IV (27ème section)
Laublet Philippe	Maître de Conférences à l'Université Paris IV (27ème section)
Seddah Djamé	Maître de Conférences à l'Université Paris IV (27ème section)

PRAG

Boldini Pascal	Docteur, PRAG à l'Université Paris IV

Ingénieurs

Le Priol Florence	Docteur, Ingénieur de recherche à l'Université Paris IV

RENTRÉE 2013 – RÉUNION D'INFORMATION

Lundi 23 septembre 2013

INSCRIPTIONS PEDAGOGIQUES
INSCRIPTIONS AUX EXAMENS

Le master se prépare en deux ans (master 1 et master 2) comportant chacun deux semestres, soit 4 semestres au total.

L'inscription administrative se fait en Sorbonne au service de la scolarité, Galerie Richelieu. Cette inscription est annuelle.

L'inscription pédagogique, qui se fait au secrétariat de l'ISHA conditionne l'inscription aux examens. Elle est donc obligatoire. Il est fortement recommandé aux étudiants de s'inscrire le plus tôt possible.

L'inscription pédagogique est semestrielle. Toutefois, il est conseillé aux étudiants de s'inscrire pour les deux semestres d'une même année au mois d'octobre.
L'inscription pour le second semestre de chaque année (S2 et S 4) pourra être modifiée selon le choix de l'étudiant (en accord avec ses enseignants).

Les inscriptions pédagogiques débuteront le lundi 23 septembre 2013.

DELIVRANCE DES ATTESTATIONS ET DES DIPLÔMES

Les étudiants ayant un diplôme complet doivent demander leur attestation dans les meilleurs délais. (Centre administratif de paris VI)
Toutefois les étudiants s'arrêtant en fin de master 1 (fin de première année) pourront recevoir le diplôme de maîtrise sur demande expresse.

Mention « *Information et communication* »

Spécialité recherche : Informatique et Ingénierie de la Langue pour la Gestion de l'Information

Responsable : Claude Montacié**,** Professeur

- Parcours **« *Logique, Sémantique, Cognition et Informatique* »,** responsable Agata Jackiewicz
- Parcours **« *Parole et Communication Homme-Machine* »** en collaboration avec l'Université René Descartes (Paris V) ; responsable : Claude Montacié

La spécialité *Informatique et Ingénierie de la Langue pour la Gestion de l'Information* (IILGI) vise à la formation de chercheurs dans les domaines de l'Ingénierie de la Langue maîtrisant la modélisation mathématique, logique, statistique et informatique d'une langue naturelle dans ses formes écrites et parlées. Les objectifs de cette formation interdisciplinaire sont, d'une part de donner à des étudiants issus des domaines des sciences du langage des compétences réelles en informatique (représentations formelles, logiques, informatiques ; utilisation d'algorithmes et de données structurées, typées …), d'autre part, de donner à des étudiants issus des domaines de l'informatique et des mathématiques, une bonne connaissance des concepts et problèmes qui relèvent des différents niveaux d'analyse linguistique (phonétique et phonologie, morphologie, syntaxe, sémantique, texte et discours, tour de parole et dialogue, …). Le but est de permettre à ces deux populations d'acquérir des connaissances en linguistique informatique : descriptions empiriques des langues compatibles avec des traitements informatiques ; Représentations logiques, statistiques et sémantiques des langues ; Constitution de ressources linguistiques.
Les deux parcours -*Logique, Sémantique, Cognition et Informatique et -Parole et Communication Homme-machine* (en collaboration avec l'université Paris Descartes) permettent d'approfondir ces connaissances et de développer des compétences pour une mise en œuvre informatique, aussi bien dans des projets théoriques que dans des applications finalisées soit en TAL (*Traitement Automatiques des Langues*), soit en TAP (*Traitement Automatique de la Parole*).
Les applications de TAL visées sont : résumé et synthèse automatiques de textes ; extraction , recherche et agrégation d'information dans les textes ; fouilles sémantiques de documents ; jugement d'évaluation et opinions ; association entre textes et images. Celles de TAP sont : opérateurs de reconnaissance et de synthèse en voix lue et spontanée ; interfaces vocales homme-machine avancées dans des contextes de dialogues finalisés ; identification de la langue et du dialecte ; extraction des caractéristiques paralinguistiques.
Cette spécialité est prise en charge par l'équipe de Linguistique Computationnelle du laboratoire Sens, Texte, Histoire et Informatique (EA 4089) de Paris Sorbonne et par l'équipe Diadex du Centre de Recherches en Informatique de Paris 5 (EA 2517).

Place dans l'offre de formation en Ingénierie de la Langue

Ce master à finalité recherche est l'un des rares à proposer une formation dans tous les domaines de l'Ingénierie de la Langue en écrit et en oral (linguistique théorique, linguistique de corpus, linguistique orale, linguistique cognitive). Tous les aspects de ces domaines sont abordés : de la modélisation théorique aux applications finales (filtrage sémantique, synthèse et reconnaissance de la parole, traduction automatique) et à leurs implications en Information et

Communication (processus de veille, documents électroniques, multilinguisme). Une autre des spécificités de ce master est une équipe enseignante à forte composante informatique (27ème section). Ces enseignants possèdent une longue expérience pédagogique dans l'enseignement de l'Informatique et du Traitement Automatique des Langues à des étudiants de Lettres et Sciences Humaines.

Compétences requises

Le master est ouvert aux étudiants de Paris-Sorbonne ayant acquis le diplôme de la licence « Langue Française et Techniques Informatique ». Il est également ouvert aux étudiants d'autres cursus (sciences du langage, langues, communication et terminologie) d'autres universités à condition que les étudiants aient acquis un certain niveau en informatique (langages formels ; une certaine habitude de la programmation et des structures de données). Il est également ouvert aux étudiants des cursus des domaines des « Sciences et Technologies » (étudiants issus des écoles d'ingénieurs ; étudiants ayant un niveau au moins égal à une licence en informatique ou en mathématiques). Un jury d'admission examine les demandes sur dossier avec, éventuellement, un entretien supplémentaire. Ce jury examinera également les demandes d'équivalence de parcours et d'unités d'enseignement.

Organisation pédagogique

La spécialité est organisée autour de 13 UE (9 communs aux 2 parcours, 2 UE spécifiques pour chaque parcours). Un parcours type est composé de 11 UE (680 heures d'enseignement) : 4 UE au 1er semestre (300 heures), 3 UE au 2ème trimestre (180 heures + préparation d'un mémoire), 3 UE au 3ème trimestre (200 heures) dont 2 UE spécifiques au parcours, 1 UE au 4ème semestre (stage et préparation d'un mémoire). La première année est consacrée à des cours théoriques sanctionnés par des examens. L'étudiant doit présenter un premier mémoire, centré sur les enseignements. Le passage en seconde année dépend des résultats obtenus en première année. Après un tronc commun sur les fondamentaux de la spécialité, la deuxième année est consacrée à des enseignements spécialisés de parcours et à un stage dans un laboratoire de recherche donnant lieu à la préparation d'un second mémoire.

Enseignements – présentation synthétique

Semestre 1	
Informatique fondamentale pour les Sciences Humaines	**6**
Représentation, acquisition et modélisation des connaissances	
Programmation objet et groupware	
Linguistique I	**6**
Epistémologie des Sciences Humaines et de l'Informatique	
Conférences de linguistique (STIH)	
Traitement automatique du langage I	***9***
Méthodologie	
Modèles de la linguistique computationnelle	
Documentation électronique	
Pluridisciplinarité et ouverture	**9**

Semestre 2	
Informatique avancée pour les Sciences Humaines	**9**
Programmation et conception objet	
Réseaux, intranet et extranet	
Plate-formes logicielles pour le TAL	
Linguistique II	**9**
Sémantique cognitive	
Linguistique de corpus et approches de l'énonciation	
Modélisations des productions orales	
Mémoire	**12**

Semestre 3 (parcours LSCI)	
Linguistique III	**9**
Perception et production de la parole	
Sciences affectives	
Jugements d'évaluation et opinions	
Logique et Cognition	**9**
Logiques et langues naturelles	
Sociolinguistique	
Web sémantique	
Traitement automatique du langage A	**12**
Recherche intelligente dans les textes	
Grammaires locales avec Unitex	
Composantes et structures textuelles pour le TAL	
Traduction automatique	

Semestre 3 (parcours PCHM)	
Linguistique III	**9**
Perception et production de la parole	
Sciences affectives	
Jugements d'évaluation et opinions	
Langage parlé et traitement de la parole	**15**
Syntaxe, sémantique et effets pragmatiques de l'oral spontané	
Reconnaissance de la parole et du locuteur	
Synthèse de la parole et émotions	
Serveurs vocaux interactifs	
Sociolinguistique	
Traitement automatique du langage B	**6**
Logiques et langues naturelles	
Traduction automatique	

Semestre 4 (parcours LSCI)	
Stage	**30**

Semestre 4 (parcours PCHM)	
Stage	**30**

1ère Année de master (M1)

Directrice des études : Agata Jackiewicz, maître de conférences

Organisation

Sur un volume horaire total de 480 heures d'enseignement pour les deux premiers semestres du parcours (18 heures par semaine en moyenne), 150 heures seront consacrées à l'enseignement de l'Informatique pour les Sciences Humaines, 240 heures à la préparation de la spécialité et 90 heures à des enseignements d'ouverture. Un projet tutoré donne lieu à un mémoire au second semestre. Les étudiants ne provenant pas de la spécialité de licence bi-disciplinaire « Langue Française et Techniques Informatiques », pourront remplacer par dérogation certains de ces EC par d'autres EC prises dans d'autres mentions ou dans les EC de niveau L3 dans le cas d'une mise à niveau.

Mutualisation

Toutes les UE de cette spécialité de M1 sont communes avec la spécialité professionnelle (parcours en Y)

Semestre 1 MIL0RS2

	Enseignements	Horaires	Modalités de contrôle des connaissances (session 1)	Crédits
UE 1	**Informatique fondamentale pour les Science Humaines M1IL100U**			**6**
EC	Représentation, acquisition et modélisation des connaissances M1IL1INF M. Laublet		1 écrit	
	Programmation objet et groupware M1IL2INF M. Montacié		1 écrit	
UE 2	**Linguistique I M1IL200U**			**6**
EC	Epistémologie des Sciences Humaines et de l'Informatique M1IL1LIN Mme Devillers		1écrit	
	Conférences de linguistique M1IL2LIN Séminaire STIH		Présence obligatoire	
UE 3	**Initiation au Traitement automatique du langage M1IL300U**			**9**
EC	Méthodologie M1IL1TAL Mme Devillers		1 écrit	
	Modèles de la linguistique computationnelle M1IL2TAL M. Seddah		1 écrit	
	Documentation électronique M1IL3TAL M. Laublet et Mme Eyharabide		1 écrit	

UE 4	**Pluridisciplinarité et ouverture M1IL400U**			**9**
3 EC à choix	EC d'un master de Linguistique française ou étrangère (autres UFR)		1 écrit	
	XML et bases de données M1IL3INF Mme Le Priol		1 écrit	
	Théories linguistiques. Approches actuelles en syntaxe et sémantique M. Bonami (L2)		1 écrit	
	Signification et théories linguistiques (L3) M. Corblin		1 écrit	
	Bases de données relationnelles (L3) L5FIBFRE Mme Jackiewicz		1 écrit	
	Grammaires formelles (L3) L5FIGRAF NN.		1 écrit	
	Mise à niveau en programmation (L1, L2 ou L3) NN.		1 écrit	

Modalités de contrôle des connaissances, session 2 (rattrapage) : épreuves orales

Semestre 2 MIL0RS2

	Enseignements	Horaires	Modalités de contrôle des connaissances (session 1)	Crédits
UE 1	**Informatique avancée pour les Sciences Humaines M2IL100U**			**9**
EC	Programmation et conception objet M2IL1INF M. Montacié		1 écrit	
	Internet, intranet et extranet M2IL2INF M. Montacié		1 écrit	
	Plate-formes logicielles pour le TAL M2IL3INF Mme Eyharabide		1 écrit	
UE 2	**Linguistiques II M2IL200U**			**9**
EC	Sémantique cognitive M2IL1IN Mme Jackiewicz		1écrit	
	Linguistique de corpus et approches de l'énonciation M2IL2LIN Mme Jackiewcz		1écrit	
EC	Modélisation des productions langagières orales M2IL3LIN M. Ibrahim		1écrit	
UE 3	**Mémoire M2IL1PRO**			**12**
	Projet tutoré et mémoire			

Modalités de contrôle des connaissances, session 2 (rattrapage) : épreuves orales

Programme des Eléments Constitutifs d'UE (EC)

Semestre 1 MIL1RS1

UE 1 - Informatique fondamentale pour les Science Humaines M1IL100U (6 ects)

a) Représentation et modélisation des connaissances M1IL1INF
Principes et systèmes de représentation des connaissances, réseaux sémantiques, logiques de description, construction de représentations conceptuelles et d'ontologies, premiers formalismes des systèmes de représentation de connaissances sur le Web (Web sémantique).
Techniques et processus d'acquisition des connaissances.

- Cours : 1h par semaine
- Travaux Dirigés/Travaux Pratiques : 1h30 par semaine

Bibliographie :
KAYSER D., *La représentation des connaissances*, Hermès, Paris, 1997, 308p.
STEFIK M., *Introduction to Knowledge Systems*, San Francisco: Morgan Kaufmann Publishers, 1993, 870 p.
ANTONIOU G. and Van HERMELEN, *A Semantic Web Primer*, The MIT Press 2004, 238.p

b) Programmation objet et groupware M1IL2INF
Paradigme objet (objet, classes et héritage),
Gestion des exceptions et des flots d'entrées-sorties
Traitement des textes et des documents XML,
Interface graphique et conception d'interfaces
Développement et test en équipe sous Eclipse (SVN., Junit)

- Cours : 1h par semaine
- Travaux Dirigés/Travaux Pratiques : 1h30 par semaine

Bibliographie:
MEYER Bertrand, *Conception et Programmation orientées objet*, Eyrolles, 2000.
BERTHIE V. et BRIAUD J.-B., *Swing la synthèse : Développement des interfaces graphiques en Java*, Dunod, 2003.
MASON M. et HURBAIN I., *Subversion : Pratique des projets collaboratifs avec SVN*, Eyrolles, 2006.

UE 2: Linguistique I M1IL200U (6 ects)

a) Epistémologie des Sciences Humaines et de l'informatique M1IL1LIN (3 ects)
Ce cours propose de traiter des concepts généraux, selon une approche épistémologique, de l'informatique et de la modélisation dans les sciences humaines. Les modèles scientifiques sont des constructions destinées à prédire certains aspects d'une réalité objective qui existe indépendamment de l'observateur. On insistera sur la différence entre les modèles, qui sont construits par les scientifiques, et la réalité, qui existe indépendamment des modèles. Les modèles sont des approximations successives de la réalité

b) Conférences de Linguistique M1IL2LIN (3 ects)
Les conférences de Linguistique sont organisées par le laboratoire Sens Texte Informatique et Histoire. Elles ont lieu salle des Actes en Sorbonne le jeudi de 18h à 20h. Le programme de l'année est disponible sur www.stih.paris-sorbonne.fr/conferences/

UE 3 –Initiation au Traitement automatique du langage M1IL300U (9 ects)

a) Méthodologie et interaction sociale M1IL1TAL
Ce cours propose une introduction aux modèles computationnels utilisés pour simuler des interactions sociales. Il introduit notamment la reconnaissance des formes et les systèmes multi-agents qui sont de plus en plus utilisés pour formaliser et étudier des problèmes d'interaction sociale. Des applications de ces modèles en recherche serviront d'illustrations. Comment valider ces modèles ? La logique de la modélisation est de confronter des données empiriques issues de l'observation aux sorties des simulations afin de valider les modèles. La problématique des bases de données de référence et des protocoles d'évaluation sera largement abordée.

b) Modèles de la linguistique computationnelle M1IL2TAL (3ects)

Le cours se propose d'introduire aux problèmes de la modélisation en syntaxe et partiellement en sémantique. Partant de l'analyse structurale de Tesnière (théorie des actants, théorie de la valence, théorie de la translation), il se poursuivra vers les théories de la dépendance et développera plus particulièrement les Grammaires Catégorielles (grammaires AB d'Adjukiewicz et Bar-Hillel) puis les Grammaires catégorielles étendues (Steedman, Biskri et Desclés). Il développera l'opposition entre les deux paradigmes : modèles syntaxiques des positions (grammaires syntagmatiques, HPSG , TAG, Grammaires fonctionnelles…) et les modèles opérateurs/ opérandes avec types fonctionnels (Harris, Shaumyan , Desclés) en développant plus particulièrement une introduction au modèle de la Grammaire Applicative et Cognitive (GA&C). Les analyses seront empruntées au français et à quelques problèmes d'autres langues non-indoeuropéennes avec des mises en œuvre informatique dans des langages de programmation fonctionnels. On insistera surtout sur les hypothèses linguistiques et sur leur discussion qui avant la formalisation et la modélisation

- Cours : 1h par semaine
- Travaux Dirigés/Travaux Pratiques : 1h30 par semaine

<u>Bibliographie</u>:

c) Documentation électronique M1IL3TAL
Ce cours a pour but de présenter différents aspects de la documentation électronique en utilisant les standards issus du monde XML sans oublier d'autres formats comme pdf.
- Formats et Normalisation des documents : les différentes solutions
- XML et DTD
- Apprendre à modéliser en XML
- Langage d'accès aux informations XPATH
- Recommandations XPointer et XLink.
- Ecriture de programmes en XSL pour le filtrage de textes.
- Systèmes de balisage et les recommandations TEI (Text Encoding Initiative).

- Cours : 1h par semaine
- Travaux Dirigés/Travaux Pratiques : 1h30 par semaine

<u>Bibliographie</u>:
HABERT B. et al, *De l'écrit au numérique : constituer, normaliser et exploiter les corpus électroniques*, InterEditions, 1998
HABERT B., *Instruments et ressources électroniques pour le français*, Ophrys 2005
HAROLD E.-R. & MEANS W.-S., *XML en concentré*, O'Reilly 2001, 2002, 2005
RAY E. T., *Introduction à XML*, O'Reilly 2001

UE 4: Pluridisciplinarité et ouverture (3 EC au choix) M1IL400U (9 ects)

a) EC du master de Linguistique ou d'un master de linguistique étrangère (se renseigner auprès des autres UFR)

b) XML et bases de données M1IL3INF

Des bases de données relationnelles à XML
Fichiers de données et DTD, Les schémas XML, Le langage de balise XSQL
Xquery : le langage d'interrogation des données XML

- Cours : 1h par semaine
- Travaux Dirigés/Travaux Pratiques : 1h30 par semaine

Bibliographie:
KATZ H., *Xquery from the experts: Guides to the W3C XML query language*, Addison-Wesley, 2003.

c) Syntaxe et sémantique (consulter la brochure de la licence LFTI en S5)
d) Linguistique française (consulter la brochure de la licence LFTI en S5)
e) Bases de données relationnelles (consulter la brochure de la licence LFTI en S5)
f) Grammaires formelles (consulter la brochure de la licence LFTI en S5)
g) Mise à niveau en programmation (consulter la brochure de la licence LFTI en S1,S3,S5)

Semestre 2 MIL1RS2

UE 1 Informatique avancé pour les Sciences Humaines M2IL100U (9 ects)

a) Programmation générique et conception objet M2IL1INF
Programmation à base de structures de données abstraites**,** polymorphisme (ad-hoc, inclusion, patrons de classe et foncteurs)
Librairie standard de composants génériques (conteneurs séquentiels et itérateurs, relation d'ordre et conteneurs associatifs, algorithmes et foncteurs, traitement de textes, expressions régulières)
Développement sous Eclipse en (débug, makefile) et interopérabilité logicielle
Méthodologies et heuristiques de conception (classes abstraites, interfaces et exceptions)

- Cours : 1h par semaine
- Travaux Dirigés/Travaux Pratiques : 1h30 par semaine

Bibliographie:
DELANNOY Claude, *Programmer en langage C++,* Eyrolles

- Cours : 1h par semaine
- Travaux Dirigés/Travaux Pratiques : 1h30 par semaine

b) Internet, intranet et extranet M2IL2INF
Réseaux de communication (Modèle OSI, Liaison de données et trames, Réseaux locaux et réseaux distants)
Programmation réseau (sockets, processus légers)
Notions de protocoles (messagerie, Langage HTML et transfert hypertexte HTTP)
Modèle de conception client-serveur (Appel de procédures à distance, Accès distant aux bases de données JDBC)
Services Web et plateforme J2EE

- Cours : 1h par semaine

- Travaux Dirigés/Travaux Pratiques : 1h30 par semaine

Bibliographie:

DJAFAAR K., *Développement J2EE avec Eclipse et WSAD*, Eyrolles, 2003.
PUJOLLE G., SALVATORI J. et NOZICK J., *Les réseaux*, Eyrolles, 2005

c) Plate-formes logicielles pour le TAL

Le but de ce module est d'introduire les principales plate-formes logicielles utilisées actuellement pour le traitement automatique des langues puis de se centrer sur l' « Architecture Générale pour le Traitement de Texte » (GATE). Après avoir présenté l'architecture de GATE, le cours s'attardera sur l'environnement de programmation graphique suivi de l'interface de programmation d'applications (API) permettant d'étendre les fonctionnalités de GATE. Partant de l'utilisation des modules parmi lesquels des analyseurs lexicaux, bases de toponymes (gazetteers), analyseurs syntaxiques, étiqueteurs ; le cours poursuivra vers des processus d'extraction d'information, comme l'annotation sémantique et l'extraction d'entités nommées. Finalement, le cours se focalisera sur l'apprentissage du langage JAPE (*Java Annotation Patterns Engine*) pour créer des règles d'annotation de documents.

- Cours : 1h par semaine
- Travaux dirigés : 1h30 par semaine

UE 2 Linguistique II M2IL200U (9 ects)

a) Sémantique cognitive M2IL1LIN

Le cours est une introduction aux courants contemporains de la sémantique cognitive et à ses méthodes. Il a pour objet de présenter (i) les origines des approches qui se réclament de la linguistique cognitive, (ii) les types de faits étudiés, (ii) les problématiques et les principaux paradigmes théoriques, (iii) les collaborations interdisciplinaires.
Le cours décrit brièvement quelques modèles (Jackendoff, Talmy, Lakoff, Pottier, Culioli…) avec leurs hypothèses de base. Il traitera de quelques problèmes linguistiques choisis : problème de la polysémie, analyse de la causalité, de la métaphore et de l'évaluation.
Enfin, on abordera la question de la place des approches cognitives dans le traitement et la modélisation informatique du langage et des langues naturelles.

- Cours : 1h par semaine
- Travaux Dirigés/Travaux Pratiques : 1h30 par semaine

Bibliographie:

b) Linguistique de corpus et approches de l'énonciation M2IL2LIN

Pour décrire les phénomènes linguistiques, il est possible désormais de disposer d'une masse considérable de données textuelles disponibles sous forme électronique, les corpus.
Le cours permet de découvrir les principaux aspects de l'analyse linguistique sur corpus, avec pour domaine d'application le champ de l'énonciation.
Approches de l'énonciation. Paramètres et indices de l'énonciation. Prise en charge. Discours rapporté. Jugement d'évaluation.
Analyse sur corpus versus analyse par introspection. Constitution de corpus. Typologie des corpus.
Vérification d'hypothèses linguistiques sur corpus. Construction et exploitation de grilles de codage.
Acquisition et organisation de ressources linguistiques à partir de corpus.
Exploitation de la dimension énonciative des textes dans des applications du TAL (résumé, fouille d'opinions…).

Frantext, Cordial, Tropes…

- Cours : 1h par semaine
- Travaux Dirigés/Travaux Pratiques : 1h30 par semaine

c) Modélisation des productions langagières orales M2IL3LIN

Synthèse des acquis concernant les différences entre les caractéristiques de l'écrit et celles de l'oral, notamment les traits pertinents pour une description et une interprétation de surface (morpho-syntaxe du français parlé).

Modélisation des régularités sonores (organisation rythmique, structuration syllabique, hésitations et reprises) qui rendent possible l'interprétation du sens d'une production orale.

Le cours s'appuiera sur un corpus d'oral spontané et posera quelques jalons élémentaires concernant la méthodologie de leur traitement expérimental.

- Cours : 1h par semaine
- Travaux Dirigés/Travaux Pratiques : 1h30 par semaine

Bibliographie:

Autour du corpus de référence du français parlé, 2004, Collection *Recherches sur le français parlé* n°18 du GARS (*Groupe aixois de recherche en syntaxe),* Recueil d'articles (Paul Cappeau, Équipe DELIC, Frédéric Sabio, Mireille Bilger, Catherine Chanet, Mylène Blasco-Dulbecco, Sandrine Caddéo, Marie-Noëlle Roubaud, Estelle Campione, Sandrine Henry & Berthille Pallaud), Publications de l'Université de Provence, 265p.

BAUDE, Olivier, 2007, *Le corpus d'Orléans,* Communication à la journée d'étude de la BNF *Autour du français parlé : de Brunot à nos jours. De l'archivage à l'exploitation*, Paris, 18 mai 2007.
Liens d'accès :
http://www.univ-orleans.fr/eslo/IMG/pdf/Le_corpus_d_Orleans_BnF_Baude_part1.pdf
http://www.univ-orleans.fr/eslo/IMG/pdf/Le_corpus_d_Orleans_BnF_Baude_part2.pdf

BLANCHE-BENVENISTE, Claire, 1997, *Approches de la langue parlée en français,* Paris: Ophrys, 164p.

DELATTRE, Pierre, 1996, Les dix Intonations de base du français, *The French Review*, Vol. 40, N° 1, pp. 1-14. Stable URL:
http://links.jstor.org/sici?sici=0016-111X%28196610%2940%3A1%3C1%3ALDIDBD%3E2.0.CO%3B2-W

UE 3: Projet tutoré et mémoire M2IL300U (12 ects)

Projet tutoré et mémoire

2ème Année de master (M2)

Organisation

Le premier semestre du M2 (18 semaines) est consacré aux enseignements fondamentaux. Le second semestre comprend des enseignements des séminaires. Il se termine par un stage obligatoire de 4 mois minimum. Le volume horaire des enseignements est 160 heures pour les deux parcours. Le passage de M1 en M2 est sélectif. Dans chaque parcours, les demandes d'inscription seront examinées par un jury composé de l'équipe pédagogique.

Parcours « *Logique, Sémantique, Cognition et Informatique* » (180 heures)

Responsable : Agata Jackiewicz

Ce parcours interdisciplinaire donne une formation fondamentale de haut niveau en linguistique théorique, cognitive et informatique conduisant à une analyse logique et mathématique orientée vers une modélisation informatique du langage appréhendé à partir des textes. Outre l'étude des grands modèles de catégorisation (grammaticale et lexicale), de sémantique cognitive et computationnelle, de syntaxe formelle, les domaines d'application sont les représentations sémantiques sous-jacentes aux bases de données lexicales informatisées multilingues, aux interfaces en langue naturelle dans les systèmes de communication homme-machine, aux recherches, extractions et synthèses d'informations, par exploration contextuelle, à partir de documents textuels, aux constructions de connaissances et d'ontologies formelles…

Les débouchés sont non seulement dans la recherche fondamentale (linguistique théorique et informatique, domaines de la cognition naturelle) mais aussi dans les secteurs de l'ingénierie linguistique (grands organismes confrontés à la gestion "intelligente" de documents textuels : résumé automatique, traduction automatique).

Semestre 3 MIL1RS4

	Enseignements	Horaires	Modalités de contrôle des connaissances (session 1)	Crédits
UE 1	**Linguistiques III M3IL110U**			**9**
	Sciences affectives : émotion, langage et cognition M3IL1LIN Mme Devillers	20 h	1écrit	
	Perception et production de la parole M3IL2LIN M. Montacié & Mme Caraty	18 h	1écrit	
	Jugements d'évaluation, opinions et sentiments M3IL3LIN Mme Jackiewicz	20 h	1 écrit	
UE 2	**Logique et Cognition M3IL320U**			**9**
EC	Logiques et langues naturelles M3IL1LOG M.Boldini	20 h	1 écrit	
	Sociolinguistique Mme Muni Tokké	20 h	1 écrit	
UE 3	**Traitement automatique du langage A M3IL330U**			**12**
EC	Composantes et structures textuelles pour le TAL M3IL3LIN Mme Jackiewicz	20 h	1 écrit	
	Grammaires locales avec Unitex M3IL1TAL Mme Kyriacopoulou	20 h	1 écrit	
	Traduction automatique M3IL4TAL M. Seddah	20 h	1 écrit	
	Recherche intelligente dans les textes M3IL1COM Mme Eyharabide	20 h	1 écrit	
	Web sémantique M3IL1INF M. Laublet	20 h	1 écrit	

Semestre 4 MIL1RS4

	Enseignements	Horaires	Modalités de contrôle des connaissances (session 1)	Crédits
UE	**Stage et mémoire M4IIL310U**			**30**
EC	Stage dans un laboratoire et mémoire de recherche **M4IL1STA**			

Modalités de contrôle des connaissances, session 2 (rattrapage) : épreuves orales

Programme des éléments constitutifs d'UE (EC)

Semestre 3

UE 1: Linguistique III M3IL110U (9 ects)

a) "Sciences affectives : émotion, langage et cognition" M3IL1LIN

La problématique des sciences affectives fait aujourd'hui un retour en force dans la communauté scientifique que ce soit en sciences humaines et sociales ou en sciences de l'information et de la communication, cela en partie grâce aux récentes découvertes en neurosciences. Les sciences humaines, la psychologie, l'éthologie et la sociologie ne sont plus les seules disciplines concernées par cette problématique : les linguistes, qu'ils soient spécialistes du discours, des interactions sociales ou de la parole, s'intéressent depuis peu aux émotions. Les émotions appellent un traitement pluridisciplinaire, qu'il s'agisse d'analyser les processus neurophysiologiques qui les sous-tendent ou leurs manifestations verbales, vocales et non verbales, qu'il s'agisse de les appréhender en laboratoire ou dans des conditions écologiques, qu'il s'agisse de les théoriser ou d'analyser leur rôle dans les processus cognitifs, les activités d'apprentissage ou les conduites de communication ou d'interaction sociale. Ce cours s'attache à montrer l'aspect pluridisciplinaire des sciences affectives et propose une introduction aux modèles computationnels utilisés pour simuler des processus affectifs en interaction sociale. Un grand nombre d'outils de simulation sera présenté à travers des réalisations en recherche pour diverses applications

b) Production et perception de la parole M3IL2LIN

Ce cours présente les mécanismes physiologiques, acoustiques et cognitifs qui sous-tendent la perception et la production des sons, des mots et des énoncés oraux de la langue parlée.
Articulatoire et acoustique (communication parlée, anatomie du cortex cérébral, phonétique et phonologie, phonématique, introduction à la phonétique articulatoire et acoustique), Prosodie (définitions et généralités, caractérisations de la substance prosodique, étiquetage prosodique, accentuation, intonation, méthodes d'analyse), Phonétique acoustique (onde acoustique, transduction, numérisation, analyse fréquentielle, modélisation de la production, analyse spectrale, coarticulation, lecture de spectrogrammes), Phonétique articulatoire, Perception auditive (psychoacoustique, sonie des sons purs, masquage fréquentiel et temporel, physiologie du système auditif, application d'un modèle psychoacoustique)

- Cours : 18 heures

c) Jugements d'évaluation, opinions et sentiments M3IL3LIN

Le développement (amplifié depuis l'apparition du Web 2.0) des échanges, des interactions entre utilisateurs, des espaces de débats aboutit à une production abondante de textes dans lesquels la subjectivité occupe une très grande place. Un nouveau champ de recherches s'est rapidement constitué autour du traitement informatisé des opinions, des attitudes, des sentiments ou des jugements d'évaluation exprimés dans divers types de documents (produits par des médias, des réseaux sociaux…).
Cette problématique renvoie à des enjeux économiques et sociaux réels, notamment en matière de veille économique, de veille d'image et de réputation, de mise en place d'observatoires politiques. D'un point de vue académique, les objectifs scientifiques sous-jacents sont multiples. Ils visent notamment (i) la modélisation linguistique et informatique des opinions, des jugements d'évaluation, des sentiments ou des émotions ; (ii) la construction, l'acquisition et la validation des ressources linguistiques (lexiques, grammaires, constructions...) pour le traitement des discours évaluatifs ; (iii) les méthodes pour identifier, annoter et extraire des opinions (des sentiments...) dans des documents textuels (audiovisuels…) ; (iv) la présentation synthétique de la diversité des données d'opinion.
Le cours propose une présentation des recherches actuelles en TAL et en linguistique sur les questions évoquées ci-dessus. Il dresse un panorama des ressources, des applications et des outils disponibles.

- Cours : 20 heures

UE 2: Logique et cognition M3IL320U (9 ects)

a) Logique et langues naturelles M3IL1LOG

Mise en place des principaux formalismes logiques pour l'analyse des langues naturelles : (a) théorie des types syntaxiques, logiques, sémantico-cognitifs (types de Church et types de Martin Löf); (b) « langages applicatifs », λ-calcul (Church), logique combinatoire (Curry et Quine), théorie de combinateurs appliquée à la linguistique et à la logique ; (c) déduction naturelle (Gentzen), β-réduction, β-expansion, signification linguistique du théorème de Church-Rosser pour la formulation de grands problèmes linguistiques (paraphrases, diathèses, thématisation, détermination, quantification, réflexivisation, anaphores, variations paradigmatiques …) ; (d) présentation des formalismes logiques sous-jacents à quelques « grands modèles linguistiques » : Chomsky (théorie des mouvements) ; Grammaires catégorielles (Calcul de Lambek, Steedman), Modélisation sémantique de Montague ; Harris (Opérateurs/opérandes ; réductions et paraphrases) ; Schaumyan (Grammaire Applicative : Langage génotype et langues phénotypes) ; « Grammaires Applicative et Cognitive » ; Melchouck (Modèle « Sens-textes » à 7 niveaux), Dick (Grammaire fonctionnelle), , … ; (e) liens avec la programmation applicative (langage CAML) et la sémantique des langages de programmation ; (f) problèmes logiques : auto-applicabilité, auto-référence, analyse logique de quelques paradoxes (Russell, Curry…).

- Cours : 20 heures

b) Sociolinguistique

Ce cours a pour objectif d'apporter aux étudiants des éléments de réflexion d'ordre épistémologique : quels sont les apports de l'approche sociolinguistique pour les sciences du langage ?
On prendra ici le terme «sociolinguistique » au sens large, puisqu'il s'agira d'examiner les diverses sous-disciplines qui incluent, à des degrés divers, une dimension sociologique dans leur cadre d'analyse. Le point d'entrée sera le concept de locuteur: puisqu'un sujet parlant est nécessairement à l'origine de tout énoncé traité en linguistique, en quoi les catégorisations qu'on lui attribue au départ influencent-elles le résultat de l'analyse ? L'arrière-plan théorique est ici celui de la ligne de partage que l'on peut construire, dans le champ des sciences du langage,

entre des linguistiques "de la langue" (les linguistiques structurales et plus spécifiquement les grammaires génératives, qui ne traitent pas directement la dimension sociale du sujet parlant) et des linguistiques "de la parole" (par exemple la linguistique interactionnelle, fortement ancrée dans des considérations sociolinguistiques).
Deux axes de travail principaux seront proposés dans cette perspective :
i) en quoi l'analyse du discours « à la française », en ce qu'elle repose sur une décontextualisation des énoncés et permet aisément un traitement quantitatif, se différencie-t-elle de la discourse analysis anglo-saxonne ?
ii) en quoi la constitution d'un corpus peut-elle être qualifiée de « sociolinguistique », «anthropologique », ou encore « écologique » ? Quelles options théoriques ces termes recouvrent-ils ? On s'intéressera particulièrement aux rapports ainsi construits entre données brutes, corpus d'étude et métadonnées.

C) Web sémantique M3IL1INF

Le web sémantique se veut un web dont le contenu peut être appréhendé et exploité par des machines. Ainsi, le web sémantique peut fournir des services plus aboutis à ses utilisateurs (trouver l'information pertinente, sélectionner, localiser et activer le service nécessaire...). Il peut être vu comme une infrastructure complémentant le contenu informel du web actuel avec de la connaissance formalisée. Il peut conduire à faire cohabiter plusieurs degrés de formalisations allant de schémas de métadonnées figées (comme celui du Dublin core) à des langages de représentation plus complexes (comme RDF, OWL et différentes logiques). Il s'appuie largement sur les ontologies. Ce cours prolonge le cours représentation de connaissances de première année et débouche sur de nombreuses applications ainsi que sur des recherches se développant rapidement au niveau mondial.

- Cours : 20 heures

UE 3: Traitement automatique du langage A M3IL330U (12 ects)

a) Composantes et structures textuelles pour le TAL M3IL3LIN

La prise en considération de l'organisation des documents textuels (structures thématique, rhétorique, énonciative, physique...) apparaît aujourd'hui comme une nécessité dans un nombre croissant d'applications informatiques : résumé ou synthèse de documents, navigation documentaire, fouille de textes à partir de points de vue, recherche d'information. On note un intérêt marqué pour les structures (ou unités) sémantiques « à gros grain » (cadres de discours, énumérations, discours rapportés...), et ainsi que pour les relations (rhétoriques...) créatrices de ces unités (approches de type RST, SDRT...). Le défi pour le domaine est d'identifier des corrélats linguistiques de fonctions discursives spécifiques, clairement identifiées, constitutives d'actes de discours de complexité variable (acte de conclure, de reformuler, de rapporter le discours d'un tiers...).
Le cours présente les principaux travaux de recherche actuels concernés par la problématique énoncée ci-dessus. Il aborde en détails le traitement des relations causales, des énumérations ainsi que celui des séquences de discours rapportés.

- Cours : 20 heures

b) Grammaires locales avec Unitex M3IL1TAL

En TAL (traitement automatique des langues), la première étape est l'analyse morphologique qui fait appel à de grands dictionnaires représentés sous la forme d'automates acycliques et de

transducteurs finis. Dans les langues à flexion, comme la plupart des langues européennes, le maximum d'efficacité est obtenu avec les dictionnaires électroniques qui comprennent toutes les formes fléchies, c'est-à-dire par exemple toutes les formes conjuguées des verbes. A l'aide du logiciel Unitex nous proposons une description formelle des unités élémentaires des langues afin d'obtenir des dictionnaires électroniques du français (mots simples et mots composés) des formes fléchies.
Par ailleurs, la description linguistique (grammaticale, morphologique, syntaxique et sémantique) est destinée à être exploitée dans des applications informatiques. Le passage de la description linguistique à la réalisation informatique est facilité par l'utilisation d'automates finis. Du point de vue de la description linguistique les automates finis, même lorsqu'ils sont utilisés pour abréger l'énumération d'ensembles finis, permettent une mise en facteur des objets sans laquelle toute description précise serait impossible. D'autre part, les données étant particulièrement volumineuses, les réalisations informatiques nécessitent des formalismes et algorithmes économes en temps. Le logiciel Unitex permet la représentation des données par automates et transducteurs qui est adaptée à cette contrainte.

COURTOIS, Blandine, « Un système de dictionnaires électroniques pour les mots simples du français », in Courtois B. et Silberztein, M. (éds), *Dictionnaires électroniques du français*, *Langue Française,* n° 87, Larousse, Paris, 1990, p. 11-22.
FRIBURGER, Nathalie, MAUREL Denis, « Finite-state transducer cascade to extract named entities in texts », *Theoretical Computer Science*, 2004, vol. 313, p. 94-104.
HOBBS, Jerry, APPELT, Douglas, BEAR, John, ISRAEL, David, KAMEYAMA, Megumi, STICKEL Mark et TYSON, Mabry, « Fastus : a cascaded finite-state transducer for extracting information from natural-language text», in ROCHE, Emmanuel, et SCHABES, Yves. (éds), *Finite State Devices for Natural Language Processing*, MIT Press, Cambridge, USA, 1996, p. 383-406.
PAUMIER, Sébastien, <www-igm.univ-mlv.fr/~unitex/manuel.htm>, 2011.

- Cours : 20 heures

c) Traduction automatique M3IL4TAL

Ce cours aborde les problèmes de la Traduction Automatique (TA) à travers un tour d'horizon des différents paradigmes qui ont ponctué son histoire. Les concepts de *langages pivot*, d'*Interlingua* ainsi que la problématique liée aux différents niveaux de transfert entre représentations morphologiques, syntaxiques, sémantiques ou pragmatiques seront présentés. On présentera la TA en mettant aussi en évidence les apports des méthodes d'apprentissage statistique dans toutes les phases *du triangle de Vauquois* en en soulignant les éventuelles limites. Une partie des séances sera consacrée à l'étude des problèmes posés par des phénomènes langagiers, syntaxiques ou sémantiques, dans un contexte de TA. La question des ressources nécessaires à la TA (étiqueteurs morpho-syntaxiques, dictionnaires informatisés, grammaires, réseaux type WordNet…) et les impacts de la TA en terme de marché économique seront aussi abordés.

- Cours : 20 heures

d) Recherche intelligente dans les textes – M3IL1COM

- **Introduction à la recherche d'information**
- **Modèles de recherche d'information** (mathématiques, cognitifs)
- **Systèmes d'organisation d'information** (Taxonomies, Thésaurus, Réseaux sémantiques, Ontologies)

- **Instruments et outils de recherche** (Annuaires, Moteurs de recherche, Moteurs spécialisés, Méta-moteurs, Méta-répertoires, Moteurs sémantiques)

- **Moteurs de recherche (**Définition, Architecture, Fonctionnement, Problématiques, Comparaison)

- **Méthodologie de recherche d'information** (Préparation, Recherche, Analyse, Évaluation, Présentation)

- **Recherche d'information dans le Web** (Indexation, L'exemple de Google, Algorithme PageRank, Référencement)

- **Stratégies de Recherche d'information** (Espace vectoriel, Modèles bayésiens, Clustering, Classement)

- Cours : 20 heures

Semestre 4 M4IL310U

UE 1 –Stage et mémoire M4IL1STA (30ects)

a) Stage dans un laboratoire et mémoire de recherche

Le mémoire de préparation à la recherche constitue l'initiation pratique à la recherche fondamentale ou appliquée. Il permet à l'étudiant d'évaluer ses motivations, de connaître et participer à la vie d'un laboratoire au sein de l'Université ou d'une grande entreprise, et l'amène à réaliser un travail personnel demandant un investissement important. À la fin de son stage dans un laboratoire, l'étudiant doit remettre un mémoire écrit et faire une présentation orale devant un jury de soutenance.

Parcours « *Parole et Communication Homme-Machine* » en collaboration avec l'Université René Descartes (Paris 5)

Responsable : Claude Montacié, professeur

L'objectif du parcours *Parole et Communication Homme-Machine* est de donner aux étudiants une formation sur les aspects fondamentaux du traitement automatique du langage parlé (les modélisations linguistiques, physiologiques et acoustiques, les paradigmes de reconnaissance et de synthèse). Cette formation, par essence pluridisciplinaire, est commune avec l'université Paris 5 (parcours de même nom de la spécialité *Informatique pour la Perception, la Cognition et les Communications* du master *Mathématiques et Informatique*) d'où proviendra une partie des étudiants. Le domaine d'application privilégié est la conception et le développement d'interfaces homme-machine avancées mettant en œuvre la reconnaissance, la compréhension, et la synthèse de la parole dans des contextes de dialogues finalisés (recherche d'information, résumés vocaux, identification de la langue et du dialecte). D'autres domaines d'applications sur la qualité de la parole comme la reconnaissance de l'identité du locuteur et de son état émotionnel sont également traités. Les débouchés sont dans la recherche fondamentale mais également dans les métiers du vocal (ingénieur technologies vocales, Ingénieur serveur vocal interactif).

Semestre 3 MIL3RS3

	Enseignements	Horaires	Modalités de contrôle des connaissances (session 1)	Crédits
UE 1	**Linguistique III M3IL110U**			**9**
3 EC	Sciences affectives : émotion, langage et cognition Mme Devillers	20 h	1 écrit	
	Perception et production de la parole M3IL5LIN M. Montacié & Mme Caraty	20 h	1écrit	
	Jugements d'évaluation, opinions et sentiments M3IL3LIN Mme Jackiewicz	20 h	1 écrit	
UE 2	**Langage parlé et traitement de la parole M3IL120U**			**15**
5 EC	Syntaxe, sémantique et effets pragmatiques de l'oral spontané M3IL7INF M. Ibrahim	20 h	1 écrit	
	Reconnaissance de la parole et du locuteur M3IL5INF Mme Caraty	20 h	1écrit	
	Synthèse de la parole et émotions M3IL4INF M. Montacié	20 h	1écrit	
	Serveurs vocaux interactifs M3IL6INF Mme Caraty	20 h	1écrit	
	Sociolinguistique Mme Muni Tokké	20 h	1 écrit	
UE 4	**Traitement automatique du langage B M3IL130U**			**6**
2 EC	Logiques et langues naturelles M.Boldini	20h	1écrit	
	Traduction automatique M3IL4TAL M. Seddah	20 h	1écrit	

Semestre 4 M4IL110L

	Enseignements	Horaires	Modalités de contrôle des connaissances (session 1)	Crédits
UE	**Stage et mémoire M4IL2STA**			**30**
EC	Stage dans un laboratoire et mémoire de recherche			

Modalités de contrôle des connaissances, session 2 (rattrapage) : épreuves orales

Programme des éléments constitutifs d'UE (EC)

Semestre 3 MIL3RS3

UE 1: Linguistique III M3IL110U (9 ects)

a) "Sciences affectives : émotion, langage et cognition" M3IL1LIN

La problématique des sciences affectives fait aujourd'hui un retour en force dans la communauté scientifique que ce soit en sciences humaines et sociales ou en sciences de l'information et de la communication, cela en partie grâce aux récentes découvertes en neurosciences. Les sciences humaines, la psychologie, l'éthologie et la sociologie ne sont plus les seules disciplines concernées par cette problématique : les linguistes, qu'ils soient spécialistes du discours, des interactions sociales ou de la parole, s'intéressent depuis peu aux émotions. Les émotions appellent un traitement pluridisciplinaire, qu'il s'agisse d'analyser les processus neurophysiologiques qui les sous-tendent ou leurs manifestations verbales, vocales et non verbales, qu'il s'agisse de les appréhender en laboratoire ou dans des conditions écologiques, qu'il s'agisse de les théoriser ou d'analyser leur rôle dans les processus cognitifs, les activités d'apprentissage ou les conduites de communication ou d'interaction sociale. Ce cours s'attache à montrer l'aspect pluridisciplinaire des sciences affectives et propose une introduction aux modèles computationnels utilisés pour simuler des processus affectifs en interaction sociale. Un grand nombre d'outils de simulation sera présenté à travers des réalisations en recherche pour diverses applications.

b) Production et perception de la parole M3IL2LIN

Ce cours présente les mécanismes physiologiques, acoustiques et cognitifs qui sous-tendent la perception et la production des sons, des mots et des énoncés oraux de la langue parlée.
Articulatoire et acoustique (communication parlée, anatomie du cortex cérébral, phonétique et phonologie, phonématique, introduction à la phonétique articulatoire et acoustique), Prosodie (définitions et généralités, caractérisations de la substance prosodique, étiquetage prosodique, accentuation, intonation, méthodes d'analyse), Phonétique acoustique (onde acoustique, transduction, numérisation, analyse fréquentielle, modélisation de la production, analyse spectrale, coarticulation, lecture de spectrogrammes), Phonétique articulatoire, Perception auditive (psychoacoustique, sonie des sons purs, masquage fréquentiel et temporel, physiologie du système auditif, application d'un modèle psychoacoustique)

- Cours : 18 heures

c) Jugements d'évaluation, opinions et sentiments M3IL3LIN

Le développement (amplifié depuis l'apparition du Web 2.0) des échanges, des interactions entre utilisateurs, des espaces de débats aboutit à une production abondante de textes dans lesquels la subjectivité occupe une très grande place. Un nouveau champ de recherches s'est rapidement constitué autour du traitement informatisé des opinions, des attitudes, des sentiments ou des jugements d'évaluation exprimés dans divers types de documents (produits par des médias, des réseaux sociaux…).
Cette problématique renvoie à des enjeux économiques et sociaux réels, notamment en matière de veille économique, de veille d'image et de réputation, de mise en place d'observatoires politiques. D'un point de vue académique, les objectifs scientifiques sous-jacents sont multiples. Ils visent notamment (i) la modélisation linguistique et informatique des opinions, des jugements d'évaluation, des sentiments ou des émotions ; (ii) la construction, l'acquisition et la validation des ressources linguistiques (lexiques, grammaires, constructions...) pour le traitement des discours évaluatifs ; (iii) les méthodes pour identifier, annoter et extraire des opinions (des sentiments...) dans des documents textuels (audiovisuels…) ; (iv) la présentation synthétique de la diversité des données d'opinion.
Le cours propose une présentation des recherches actuelles en TAL et en linguistique sur les questions évoquées ci-dessus. Il dresse un panorama des ressources, des applications et des outils disponibles.

- Cours : 20 heures

UE 2: Langage parlé et traitement de la parole M3IL120L (15 ects)

a) Syntaxe, sémantique et effets pragmatiques de l'oral spontané M3IL7INF

Ce cours présente les caractéristiques syntaxiques, sémantiques et pragmatiques qui obligent à traiter spécifiquement l'oral spontané.
L'oral spontané se distingue radicalement de l'écrit du fait qu'il implique toujours – à l'exception des monologues qui constituent une irruption de l'écrit dans l'oral – un discours partagé avec pour conséquence la plus immédiate que la reconstruction des séquences effacées, interrompues ou tronquées dans le discours d'un locuteur dépend à la fois des discours de ses interlocuteurs et de paramètres relativement bien typés de l'environnement discursif. D'autre part, contrairement à l'écrit qui renvoie toujours à un faire différé, l'oral spontané s'articule en permanence à un faire potentiellement immédiat dont les protocoles d'accomplissement sont très rigoureusement balisés par la langue. L'oral spontané a de ce fait un mode spécifique de production de l'information et de construction de sens dont il convient, en prévision d'un traitement automatisé, de bien connaître les mécanismes afin de réussir la reconstruction de séquences indispensables à la cohérence des chaînes sonores analysées.

- Cours : 20 heures

b) Synthèse de la parole et émotions M3IL4INF
Ce cours présente les paradigmes de la synthèse de la parole à partir du texte, de l'analyse synthèse et des algorithmes associés, les effets de facteurs physiologiques, pathologiques, psychologiques et comportementaux sur la parole.
Synthèse à partir du texte (conversion graphèmes-phonèmes et interface syntaxe-prosodie)
Prédiction linéaire (modèle linéaire source/filtre, méthode d'autocorrélation, vocodeur)
Codage de la parole, analyse-synthèse et synthèse acoustique
Algorithmes de synthèse par corpus
Voix pathologiques (dysphonies, dysarthries, dysprosodie) et qualité de la parole (caractérisation)
Emotions et parole émotionnelle (prosodie de l'émotion, détection de l'état émotionnel)
Stress et parole stressée (indicateurs du stress et de la fatigue, analyse de la voix stressée

- Cours : 20 heures

c) Reconnaissance de la parole et du locuteur M3IL5INF

Ce cours présente les paradigmes de la reconnaissance de la parole et des caractéristiques individuelles du locuteur et les algorithmes associés.
Reconnaissance des formes vocales (tâches et objectifs en reconnaissance, principe des applications, espace de représentation des formes, concept de classe, apprentissage des classes, reconnaissance/identification), Techniques de reconnaissance des formes vocales (stratégies de reconnaissance, fusion des classificateurs, adaptation par maximum a posteriori), Tâches de reconnaissance (facteurs de complexité du signal vocal, facteurs de complexité de la tâche de reconnaissance), Reconnaissance vocale (reconnaissance de continuums acoustiques, programmation dynamique, comparaison dynamique, application à la reconnaissance d'entités vocales isolées), modèles de Markov cachés (théorie de l'information, processus de Markov, modélisation du canal acoustique, méthodes d'apprentissage, application à la reconnaissance en mots isolés et en parole continue). Biométrie physique et caractéristiques individuelles (authentification auditive et visuelle), Reconnaissance du locuteur (identification, vérification, suivi, détection de changement)

- Cours : 20 heures

d) Serveurs vocaux interactifs M3IL6INF

Ce cours a pour but la conception et mise au point d'un dialogue oral finalisé et son déploiement sur un serveur vocal interactif.
Modèle de tâche (GOMS, CTT) et modèle de dialogue (Réseaux de Petri à objets, …)
Architecture de dialogue (ARCH, AMF) et processus de conception et d'adaptation (Plastic ML)
Structures de dialogues et flux de commande en Voice XML
Ressources, entrées utilisateur et sorties du système en Voice XML
Mise en œuvre de solutions PABX open-source (Asterix)

- Cours : 20 heures

e) Sociolinguistique

Eléments de réflexion en sociolinguistique
Ce cours a pour objectif d'apporter aux étudiants des éléments de réflexion d'ordre épistémologique : quels sont les apports de l'approche sociolinguistique pour les sciences du langage ?
On prendra ici le terme «sociolinguistique » au sens large, puisqu'il s'agira d'examiner les diverses sous-disciplines qui incluent, à des degrés divers, une dimension sociologique dans leur cadre d'analyse. Le point d'entrée sera le concept de locuteur: puisqu'un sujet parlant est nécessairement à l'origine de tout énoncé traité en linguistique, en quoi les catégorisations qu'on lui attribue au départ influencent-elles le résultat de l'analyse ? L'arrière-plan théorique est ici celui de la ligne de partage que l'on peut construire, dans le champ des sciences du langage, entre des linguistiques "de la langue" (les linguistiques structurales et plus spécifiquement les grammaires génératives, qui ne traitent pas directement la dimension sociale du sujet parlant) et des linguistiques "de la parole" (par exemple la linguistique interactionnelle, fortement ancrée dans des considérations sociolinguistiques).
Deux axes de travail principaux seront proposés dans cette perspective :
i) en quoi l'analyse du discours « à la française », en ce qu'elle repose sur une décontextualisation des énoncés et permet aisément un traitement quantitatif, se différencie-t-elle de la discourse analysis anglo-saxonne ?
ii) en quoi la constitution d'un corpus peut-elle être qualifiée de « sociolinguistique », «anthropologique », ou encore « écologique » ? Quelles options théoriques ces termes recouvrent-ils ? On s'intéressera particulièrement aux rapports ainsi construits entre données brutes, corpus d'étude et métadonnées.

UE 3 Traitement automatique du langage B M3IL130U (6 ects)

a) Logique et langues naturelles M3IL1LOG

Mise en place des principaux formalismes logiques pour l'analyse des langues naturelles : (a) théorie des types syntaxiques, logiques, sémantico-cognitifs (types de Church et types de Martin Löf); (b) « langages applicatifs », λ-calcul (Church), logique combinatoire (Curry et Quine), théorie de combinateurs appliquée à la linguistique et à la logique ; (c) déduction naturelle (Gentzen), β-réduction, β-expansion, signification linguistique du théorème de Church-Rosser pour la formulation de grands problèmes linguistiques (paraphrases, diathèses, thématisation, détermination, quantification, réflexivisation, anaphores, variations paradigmatiques …) ; (d) présentation des formalismes logiques sous-jacents à quelques « grands modèles linguistiques » : Chomsky (théorie des mouvements) ; Grammaires catégorielles (Calcul de Lambek, Steedman), Modélisation sémantique de Montague ; Harris (Opérateurs/opérandes ; réductions et paraphrases) ; Schaumyan (Grammaire Applicative : Langage génotype et langues phénotypes) ; « Grammaires Applicative et Cognitive » ; Melchouck (Modèle « Sens-textes » à 7 niveaux), Dick (Grammaire fonctionnelle), , … ; (e) liens avec la programmation applicative (langage CAML) et la sémantique des langages de programmation ; (f) problèmes logiques : auto-applicabilité, auto-référence, analyse logique de quelques paradoxes (Russell, Curry…).

- **Cours : 20 heures**

b) Traduction automatique M3IL4TAL

Ce cours aborde les problèmes de la Traduction Automatique (TA) à travers un tour d'horizon des différents paradigmes qui ont ponctué son histoire. Les concepts de *langages pivot*, d'*Interlingua* ainsi que la problématique liée aux différents niveaux de transfert entre représentations morphologiques, syntaxiques, sémantiques ou pragmatiques seront présentés. On présentera la TA en mettant aussi en évidence les apports des méthodes d'apprentissage statistique dans toutes les phases *du triangle de Vauquois* en en soulignant les éventuelles limites. Une partie des séances sera consacrée à l'étude des problèmes posés par des phénomènes langagiers, syntaxiques ou sémantiques, dans un contexte de TA. La question des ressources nécessaires à la TA (étiqueteurs morpho-syntaxiques, dictionnaires informatisés, grammaires, réseaux type WordNet…) et les impacts de la TA en terme de marché économique seront aussi abordés.

- Cours : 20 heures

Semestre 4 MIL3RS4

UE 1: Stage et mémoire M4IL110U (30ects)

a) Stage dans un laboratoire et mémoire de recherche

Le mémoire de préparation à la recherche constitue l'initiation pratique à la recherche fondamentale ou appliquée. Il permet à l'étudiant d'évaluer ses motivations, de connaître et participer à la vie d'un laboratoire au sein de l'Université ou d'une grande entreprise, et l'amène à réaliser un travail personnel demandant un investissement important. À la fin de son stage dans un laboratoire, l'étudiant doit remettre un mémoire écrit et faire une présentation orale devant un jury de soutenance.



Cahier commun aux brochures de Master

Année 2013-2014

Sommaire :

DISPOSITIONS GÉNÉRALES

L'ORGANISATION DE L'ANNÉE UNIVERSITAIRE 2013-2014

L'année universitaire se divise en 2 semestres différenciés comportant chacun 13 semaines d'enseignement. Les étudiants suivent les enseignements de chacun des semestres et sont soumis à un contrôle des connaissances semestriel. Une inscription administrative se prend pour toute l'année universitaire.

Rappel : le Master s'obtient à l'issue de deux années, M1 et M2, de 2 semestres chacune.

Les cours des semestres impairs commenceront le *lundi 23 septembre 2013* et s'achèveront le *samedi 11 janvier 2014*.
Les cours des semestres pairs commenceront le *lundi 3 février* et s'achèveront le *samedi 17 mai 2014*.
Les examens relatifs aux semestres impairs auront lieu *du 13 au 25 janvier 2014*
Les examens relatifs aux semestres pairs auront lieu *du 19 au 2 juin 2014.*
Les dates de la session de rattrapage auront lieu du 26 juin au 2 juillet 2014.

LES INSCRIPTIONS

1) Inscription administrative (IA)

Elle est annuelle et obligatoire. Date de clôture : le **11 octobre 2013**

Pour plus d'information, consulter notre site :
http://www.paris-sorbonne.fr/nos-formations/comment-s-inscrire

A votre service, une adresse électronique pour répondre à vos questions concernant votre scolarité à Paris-Sorbonne (Paris IV) :

scolarite@paris-sorbonne.fr

2) Connection à l'ENT

Après avoir procédé à votre inscription administrative (et au paiement de vos droits de scolarité), connectez vous à l'ENT étudiant afin de vous créer un compte (voir le mode d'emploi sur le document ENT)

3) Inscriptions pédagogiques (IP)
Les inscriptions pédagogiques permettent de suivre les T.D. et servent d'inscription au contrôle des connaissances. **Elles sont obligatoires.**

- Pour les masters à finalité professionnelle, elles s'effectuent en ligne depuis l'ENT Etudiant (ent.paris-sorbonne.fr / Onglet « Scolarité » / IP) :

En septembre 2013 pour les semestres impairs :
du Mercredi 18 septembre 2013 à 16h au vendredi 27 septembre 2013 17h

En janvier 2014 pour les semestres pairs (les dates seront précisées sur l'ENT étudiant).

- Pour les masters à finalité recherche, elles s'effectuent dans l'UFR aux dates et heures indiquées sur son site.

LES EXAMENS

Les épreuves sont organisées par les UFR.

Les résultats seront consultables environ 3 jours après les jurys **sur l'ENT étudiant** / Scolarité / Mon dossier administratif / Notes et résultats/

Le service des examens délivre les relevés de notes semestriels sur demande des étudiants inscrits en Master.

Le service des examens est ouvert de 9h à 12h : site Sorbonne, escalier F, 2e étage, salle F355
Courriel : secretariat.examens@paris-sorbonne.fr

LES DIPLOMES

Les diplômes sont délivrés par le service des diplômes à la demande de l'étudiant.

Les formulaires de demande sont téléchargeables sur site internet www.paris-sorbonne.fr puis **Nos formations** puis **Votre scolarité** ou sur l'**ENT étudiant** (www.ent.paris-sorbonne.fr / Scolarité / Diplômes)



CALENDRIER UNIVERSITAIRE 2013-2014 voté au CA du 26 avril 2013

concerne les cursus Licence, Master et les Diplômes d'Université

1er semestre : lundi 23 septembre 2013 au samedi 1er février 2014

2è semestre : lundi 3 février au 02 juillet 2014

Début des Inscriptions Administratives : 08 juillet 2013

Date limite d'inscription administrative : 04 octobre 2013 pour les cursus LICENCE et DU
11 octobre 2013 pour le cursus MASTER

Date limite d'inscription pédagogique : 27 septembre 2013 pour les cursus Licence et Diplôme d'université, 25 octobre 2013 pour le cursus Master.

2013								2014																	
septembre 2013		octobre		novembre		décembre		janvier		février		mars		avril		mai		juin		juillet		août		septembre 2014	
1	D	1	M	1	V	1	D	1	M	1	S	1	S	1	M	1	J	1	D	1	M	1	V	1	L
2	L	2	M	2	S	2	L	2	J	2	D	2	D	2	M	2	V	2	L	2	M	2	S	2	M
3	M	3	J	3	D	3	M	3	V	3	L	3	L	3	J	3	S	3	M	3	J	3	D	3	M
4	M	4	V	4	L	4	M	4	S	4	M	4	M	4	V	4	D	4	M	4	V	4	L	4	J
5	J	5	S	5	M	5	J	5	D	5	M	5	M	5	S	5	L	5	J	5	S	5	M	5	V
6	V	6	D	6	M	6	V	6	L	6	J	6	J	6	D	6	M	6	V	6	D	6	M	6	S
7	S	7	L	7	J	7	S	7	M	7	V	7	V	7	L	7	M	7	S	7	L	7	J	7	D
8	D	8	M	8	V	8	D	8	M	8	S	8	S	8	M	8	J	8	D	8	M	8	V	8	L
9	L	9	M	9	S	9	L	9	J	9	D	9	D	9	M	9	V	9	L	9	M	9	S	9	M
10	M	10	J	10	D	10	M	10	V	10	L	10	L	10	J	10	S	10	M	10	J	10	D	10	M
11	M	11	V	11	L	11	M	11	S	11	M	11	M	11	V	11	D	11	M	11	V	11	L	11	J
12	J	12	S	12	M	12	J	12	D	12	M	12	M	12	S	12	L	12	J	12	S	12	M	12	V
13	V	13	D	13	M	13	V	13	L	13	J	13	J	13	D	13	M	13	V	13	D	13	M	13	S
14	S	14	L	14	J	14	S	14	M	14	V	14	V	14	L	14	M	14	S	14	L	14	J	14	D
15	D	15	M	15	V	15	D	15	M	15	S	15	S	15	M	15	J	15	D	15	M	15	V	15	L
16	L	16	M	16	S	16	L	16	J	16	D	16	D	16	M	16	V	16	L	16	M	16	S	16	M
17	M	17	J	17	D	17	M	17	V	17	L	17	L	17	J	17	S	17	M	17	J	17	D	17	M
18	M	18	V	18	L	18	M	18	S	18	M	18	M	18	V	18	D	18	M	18	V	18	L	18	J
19	J	19	S	19	M	19	J	19	D	19	M	19	M	19	S	19	L	19	J	19	S	19	M	19	V
20	V	20	D	20	M	20	V	20	L	20	J	20	J	20	D	20	M	20	V	20	D	20	M	20	S
21	S	21	L	21	J	21	S	21	M	21	V	21	V	21	L	21	M	21	S	21	L	21	J	21	D
22	D	22	M	22	V	22	D	22	M	22	S	22	S	22	M	22	J	22	D	22	M	22	V	22	L
23	L	23	M	23	S	23	L	23	J	23	D	23	D	23	M	23	V	23	L	23	M	23	S	23	M
24	M	24	J	24	D	24	M	24	V	24	L	24	L	24	J	24	S	24	M	24	J	24	D	24	M
25	M	25	V	25	L	25	M	25	S	25	M	25	M	25	V	25	D	25	M	25	V	25	L	25	J
26	J	26	S	26	M	26	J	26	D	26	M	26	M	26	S	26	L	26	J	26	S	26	M	26	V
27	V	27	D	27	M	27	V	27	L	27	J	27	J	27	D	27	M	27	V	27	D	27	M	27	S
28	S	28	L	28	J	28	S	28	M	28	V	28	V	28	L	28	M	28	S	28	L	28	J	28	D
29	D	29	M	29	V	29	D	29	M			29	S	29	M	29	J	29	D	29	M	29	V	29	L
30	L	30	M	30	S	30	L	30	J			30	D	30	M	30	V	30	L	30	M	30	S	30	M
		31	J			31	M	31	V			31	L			31	S			31	J	31	D		

Période de cours

Période de révision

Examens
Jurys L
Jurys M

publication résultats

Dimanches et jours fériés

Vacances universitaires

jury M rech :	1er S : 31 janvier	Jury M pro :	1er S : 31 janvier
	2èS : 08 juillet		2èS : 06 juin
	rattrapage: 26 sept		rattrapage: 08 juillet